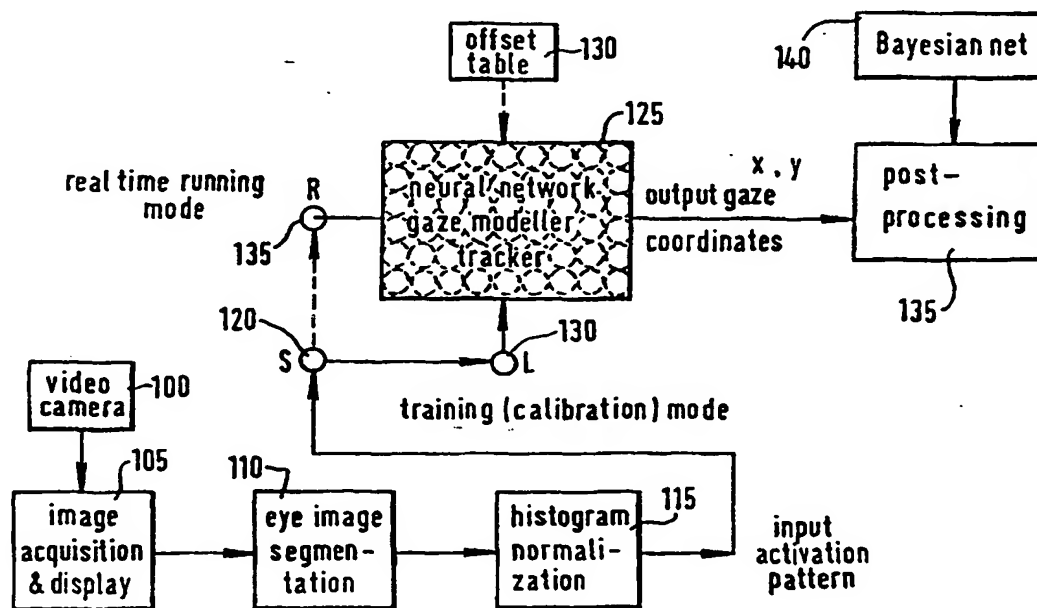




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>G06F 3/00</b>		A1	(11) International Publication Number: <b>WO 99/26126</b>
			(43) International Publication Date: 27 May 1999 (27.05.99)
(21) International Application Number: PCT/GB98/03441 (22) International Filing Date: 16 November 1998 (16.11.98) (30) Priority Data: 9724277.0                      17 November 1997 (17.11.97)      GB 98306261.3                    5 August 1998 (05.08.98)        EP (71) Applicant (for all designated States except US): BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): AZVINE, Behnam [GB/GB]; 6 Dodson Vale, Kesgrave, Ipswich, Suffolk IP5 2GT (GB). DJIAN, David [FR/GB]; Uplees, Meeting Lane, Grundisburgh, Woodbridge, Ipswich, Suffolk IP13 6UB (GB). TSUI, Kwok, Ching [GB/GB]; 138 Berkley Close, Highwoods, Colchester, Essex CO4 4RR (GB). XU, Li-Qun [GB/GB]; 34 Dodson Vale, Kesgrave, Ipswich, Suffolk IP5 2GT (GB). (74) Agents: READ, Matthew, Charles et al.; Venner, Shipley & Co., 20 Little Britain, London EC1A 7DH (GB).			(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  Published With international search report.

(54) Title: USER INTERFACE



## (57) Abstract.

A gaze tracker for a multimodal user interface uses a standard videoconferencing set on a workstation to determine where a user is looking on a screen. The gaze tracker uses the video camera (100) to make a quantised image of the user's eye. The pupil is detected in the quantised image and a neural net (125) is used in training the gaze tracker to detect gaze direction. A pre-processor (115) may be used to improve the input to the neural net. A Bayesian net (140) is used to learn the relationship between response time and accuracy for the output of the neural net so that a user's externally set preference can be accommodated.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## USER INTERFACE

## Field of the invention

The present invention relates to a user interface for a data or other software system, which monitors an eye of the user, such as a gaze tracker. The interface finds particular but not exclusive application in a multimodal-system.

## Background

- Gaze tracking is a challenging and interesting task traversing several disciplines including machine vision, cognitive science and human computer interactions (Velichkovský, B.M. and J.P. Hansen (1996): "New technological windows into mind: There is more in eyes and brains for human-computer interaction". Technical Report: Unit of Applied Cognitive Research, Dresden University of Technology, Germany). The idea that a human subject's attention and interest on a certain object, reflected implicitly by eye movements, can be captured and learned by a machine which can then act automatically on the subject's behalf lends itself to many applications, including for instance video conferencing (Yang, J., L. Wu and A. Waibel (1996): "Focus of attention in video conferencing": Technical Report, CMU-CS-96-150, School of Computer Science, Carnegie Mellon University, June 1996). This idea can be used for instance for:
- focusing on interesting objects and transmitting selected images of them through the communication networks,
  - design of a new generation of interfaces for computers to reach more users (as disclosed in Jacob, R.J.K. (1995): "Eye tracking in advanced interface design" in W. Barfield and T. Furness (eds.) - Advanced Interface Design and Virtual Environments, published by Oxford University Press, and in Nielsen, Jakob (1993): "Noncommand user interfaces" - Communications of the ACM, 36 (4), 83 - 99), and
  - the study of human vision, cognition, and attentional processes (Zangemeister, W.H., H.S. Stiehl, C. Freska (eds) (1996) - Visual Attention and Cognition published by North-Holland: Elsevier Science B.V.: Amsterdam).

Traditionally, gaze tracking uses the so-called pupil-center/corneal-reflection method (Cleveland, D. and N. Cleveland (1992)- "Eyegaze eyetracking system" Proc. of 11<sup>th</sup> Monte-Carlo International Forum on New Images, Monte-

Carlo, January 1992). This uses controlled infra-red lighting to illuminate the eye, computing the distance between the pupil centre (the bright-eye effect) and the small very bright reflection off the surface of the eye's cornea to find the line of sight on the display screen, through geometric projections. This kind of method normally involves a specialised high speed/high resolution camera, a controlled lighting source and electronic hardware equipment, and is sometimes intrusive (Stampe, D. (1993): "Heuristic filtering and reliable calibration methods for video based pupil-tracking systems"- Behaviour Research Methods, Instruments, Computers, 25 (2), pp. 137-142). The user is often requested to remain motionless during the course of operation. As a result, the gaze trackers are mostly used in constrained laboratory environments for passively capturing, recording, and playing back later overlaid time-stamped eye movement trajectories for analysis of fixation and saccade phenomena in connection with various psychophysical experimental tasks.

Recently increasing demand on intelligent systems, however, has generated a need for more convenient, effective and natural ways of communication between humans and computers. This has required expansion of the narrow-bandwidth channel from user-to-computer that is currently operated for instance through a (low speed) mouse and keyboard. Accurate extraction of eye movement information, along with speech, gestures (Darrell, T.J. and A.P. Pentland (1994): "Recognition of space-time gestures using a distributed representation" Mammone, R.J (ed.)) and other avenues, and the wise utilisation of it, have been recognised as potentially playing a part in forming a fast and natural interface, with the ability to respond actively to the user's natural viewing intention. Relevant work is published in, for example, Starker, I., R.A. Bolt (1990): "A Gaze-responsive self-disclosing display" - ACM CHI'90 Conference Proceedings; Human Factors in Computing Systems, Seattle, Washington, pp. 3-9 and in Hansen, J.P., A.W. Andersen, and P. Roed. (1995): "Eye-gaze control of multimedia systems" in Y. Anzai, K. Ogawa and H. Mori (eds): Symbiosis of Human and Artifact published by Elsevier Science.

A gaze tracker is described in US-A-5 481 622, which comprises a helmet worn by a user, which uses a camera to acquire a video image of the pupil, mounted on the helmet. A frame grabber is coupled to the camera to accept and

convert analog data from the camera into digital pixel data. A computer coupled to the frame grabber processes the digital pixel data to determine the position of the pupil. A display screen is coupled to the computer and is mounted on the helmet. The system is calibrated by the user following a cursor on the display screen while the system measures the pupil position for known locations of the cursor. However, the arrangement is complicated and requires special hardware namely the helmet arrangement and is not suited to everyday commercial use.

US-A-5 471 542 discloses a gaze tracker in which a video camera is provided on a personal computer in order to detect eye movement to perform functions similar to those achieved with a conventional hand-held mouse.

The present invention provides an improved arrangement which can be trained in order take account of characteristics of the user and the user's preferences.

## 15 Summary of the invention

According to the present invention, there is provided a user interface, for use in making inputs to a data or communications system, responsive to the user's eye, comprising:

- i) a scanning device for capturing a quantised image of an eye;
- 20 ii) a pupil image detector to detect a representation of the pupil of the eye in the quantised image;
- iii) a display for a plurality of visual targets;
- iv) a first learning device to relate at least one variable characteristic of said image of the eye to a selected one of said visual targets; and
- 25 v) a second learning device, for relating external parameters apparent to a user of the system to parameters internal to the system.

The invention also provides in another aspect a method of training the user interface, which involves displaying training data on the display and training the first learning device to relate the variable characteristic of the image of the eye to the training data when the user gazes at the displayed training data.

The invention may also include training the second learning device to relate external parameters apparent to the user of the system to the internal parameters.

The internal parameters may be a function of fixation of the gaze of the user at a particular region on the display, and the external parameters may include the

time taken to determine that a fixation has occurred and the positional accuracy thereof.

Embodiments of the present invention provide a real-time non-intrusive gaze tracking system; that is, a system which can tell where a user is looking, for instance on a computer screen. The gaze tracking system can provide a vision component of a multimodal intelligent interface, particularly suited for resolving ambiguities and tracking contextual dialogue information. However, it is also an effective piece of technology in its own right, leading to many potential applications in human-computer interactions where the ability to find human attention is of significant interest.

Robust segmentation of eye images and efficient training (calibration) of a large neural network can be provided.

Embodiments of the present invention can provide a flexible, cheap, and adequately fast gaze tracker, using a standard videoconferencing camera sitting on a workstation and without resorting to any additional hardware and special lighting. These embodiments provide a neural network based, real-time, non-intrusive gaze tracker.

Data preprocessing means may be provided to enhance the output of the scanning device for use by the learning device. For instance, where the quantised image of the eye comprises an array of pixels with associated contrast information, such data preprocessing means may comprise means to normalise said array and to allocate to each individual pixel thereof a contrast value selected from a set of discrete contrast values. By using a relatively small set of discrete contrast values, this can make the output of a standard video camera viable for processing by the learning device. Otherwise, far too much data is likely to be involved to allow the interface to be practicable.

The second learning device may be provided to relate parameters apparent to a user of the system to parameters internal to the system. This can be used to provide an adjustment capability such that the user can adjust parameters apparent in use of the system by inputting parameter information to the system, the system responding thereto by adjusting parameters internal to the system, in accordance with one or more learned relationships therebetween.

Stated generally, the invention provides a gaze tracker including means for determining when a user achieves a gaze fixation on a target, comprising learning means for learning a relationship between response time and accuracy for achieving a fixation, and means responsive to a user's preference concerning the relationship for controlling signification of the fixation. The learning means may  
5 comprise a Bayesian net.

The invention also includes a user interface for a computer workstation usable for videoconferencing, the interface being configured for use in making inputs to a  
10 data or communications system in response to movements of the user's eye, comprising:

- i) a tv videoconferencing camera to be mounted on the workstation for capturing a quantised image of an eye;
  - ii) a pupil image detector to detect a representation of the pupil of the eye in  
15 the quantised image;
  - iii) a workstation display for a plurality of visual targets; and
- a neural net to relate at least one variable characteristic of said image of the eye to a selected one of said visual targets.

## 20 Brief description of the drawings

A gaze tracker according to an embodiment of the present invention will now be described, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 shows in schematic outline a neural network based gaze  
25 modelling/tracking system as an embodiment of the present invention wherein Figure 1A illustrates the physical configuration and Figure 1B illustrates the system in terms of functional blocks;

Figure 2 shows a snapshot of a captured image of a user's head image for use in the system shown in Figure 1;

30 Figure 3 shows an example of a fully segmented eye image for use in the system shown in Figure 1;

Figure 4 shows a histogram of a segmented grayscale eye image;

Figure 5 shows a transfer function for the normalisation of segmented eye image data;

Figure 6 shows a normalised histogram version of the eye image of Figure 3;

Figure 7 shows a neural network architecture for use in the system of Figure 5 1;

Figure 8 shows a matrix of grids laid over a display screen for the collection of training data for use in a system according to Figure 1;

Figure 9 shows a Gaussian shaped output activation pattern corresponding to the vertical position of a gaze point;

10 Figure 10 shows training errors versus number of training epochs in a typical training trial of the network shown in Figure 7;

Figure 11 shows learning and validation errors versus number of training epochs in a trial of the learning process for the network shown in Figure 7; and

Figure 12 shows a histogram of the neural network's connection weights 15 after 100 training epochs.

#### Detailed description

A goal of the gaze tracker is to determine where the user is looking, within 20 the boundary of a computer display, by the appearance of eye images detected by a monitoring camera. Figure 1A shows an example of the physical configuration of the gaze tracker. A video camera 100 of the kind used for video conferencing is mounted on the display screen 101 of a computer workstation W in order to detect an eye of a user 102. The workstation includes a conventional processor 25 103 and keyboard 104. The task performed by the gaze tracker can be considered as a simulated forward-pass mapping process from a segmented eye image space, to a predefined coordinate space such as the grid matrix shown in Figure 2. The mapping function in general however is a nonlinear and highly variable one because of a variety of uncertain factors such as changes in lighting, 30 head movement and background objects moving, to name but a few.



### Methodology and system

In this section are described a methodology and system of using a "feedforward" neural network for modelling the above mentioned mapping process for gaze tracking, explaining the key techniques used for each component of the system.

There are two primary observations (constraints) underlying the method and the gaze tracking system of embodiments of the present invention described hereinafter, these being as follows:

- i) first, in a close contact (for example the normal distance of a user facing a computer screen) the appearance of an eye, in the view of an observer (a camera), is informative enough to indicate where the user is looking; and
- ii) this information is less ambiguous and easier to extract if the user's head orientation generally conforms to the line of sight of his or her eyes.

The former has been determined by experiment while the latter is introduced to avoid the unnecessary many-to-one mapping situation where a person can view an object on a screen from various head orientations.

Even when these two points are borne in mind, the actual mapping function between an eye appearance and its corresponding gaze point is still highly nonlinear and very complicated. This complexity arises from uncertainties and noise encountered at every processing/modelling stage. In particular, for instance, it can arise from errors in eye segmentation, the user's head movement, changes of the eye image depth relative to the camera, decorations around the eye, such as glasses or pencilled eyebrows, and changes in ambient lighting conditions.

Referring to Figure 1A, embodiments of the present invention work in an office environment with the simple video camera 100 mounted on the right side of the display screen 101 of the workstation W, to monitor the user's face continuously. There is no specialised hardware, such as a lighting source, involved. The user sits comfortably at a distance of about 22 to 25 inches away from the screen. He is allowed to move his head freely while looking at the screen, but needs to keep it within the field of view of the camera, and to keep his face within a search window overlaid on the image.

As shown in Figure 1B, the neural network based gaze tracker takes the output of an ordinary video camera 100, as might be used for video conferencing, and feeds it to the following functional processing blocks:

- an image acquisition and display unit 105
- 5 • an eye image segmentation unit 110
- a histogram normalisation unit 115
- a switch 120
- a neural network gaze modeller 125.

The switch 120 takes the output of the histogram normalisation unit 115  
10 and feeds it to a learning node 130 when the modeller 125 is in training mode, or to a real time running node 135 when the modeller 125 has been trained and is to be used for detecting gaze co-ordinates.

The techniques used and the functions of each processing stage are now described below.

15

#### *Image acquisition*

The analogue video signal from a low cost video camera 100 is captured and digitised by the image acquisition and display unit 105 using the SunVideo Card, a video capture and compression card for Sun SPARCstations, and the XIL imaging  
20 foundation library developed by SunSoft. (The library is described in Pratt, W.K. (1997): "Developing Visual Applications: XIL—An Imaging Foundation Library" - published by Sun Microsystems Press). XIL is a cross-platform C functions library that supports a range of video and imaging requirements. For the purpose of simplicity, only grayscale images are used in embodiments of the present invention  
25 described herein. Colour images may however be used in enhancements of the system as colours contain some unique features that are otherwise not available from grayscale images, as is shown in the recent work of Oliver and Pentland (1997), published in "LAFTER: Lips and face real time tracker" - Proc. of Computer Vision and Pattern Recognition Conference, CVPR'97. June 1997, Puerto Rico.

30 The device image of the SunVideo Card, which is a 3-banded 8-bit image in YUV colorspace, sized 768 x 576 pixels for PAL, is converted into an 8-bit grayscale image and scaled to the size of 192 pixels in width and 144 pixels in

height, in the field of view of the video camera 100. The maximum capture rate of the SunVideo Card is 25 fps for PAL.

Figure 2 shows, as an example, a snapshot of the captured user's head image in an open-plan office environment under normal illumination. It shows a head image 200 of 192 by 144 pixels and a search window 205 within the head image of 100 by 60 pixels.

### *Eye image segmentation*

The objective of this processing, done by the eye image segmentation unit 110, is first to detect the small darkest region in the pupil of the eye, and then go on to segment the proper eye image.

For this purpose, the fixed search window 205 shown in Figure 2, is started in the centre part of the grabbed image 200. Inside this search window 205, the image 200 is iteratively thresholded, initially with a lower threshold  $T_0$ . (A similar approach was adopted in a gaze tracking task of different purpose, published by Stiefelhagen, R., J. Yang, and A. Waibel (1996): "Gaze tracking for multimodal human-computer interaction", Proc. of IEEE Joint Symposia on Intelligence and Systems.)

Morphological filters (dilation and erosion) are used to remove noise or fill "gaps" of the generated binary image which is then searched pixel by pixel from top left to bottom right. Individual objects, comprising pixel clusters, are found and labelled using the 4-connectivity algorithm described in Jain, R., R. Kasturi, and B.G. Schunck (1995): "Machine Vision", published by McGraw-Hill and MIT Press. A rectangular blob is used to represent each found object. Unless a reasonable number of objects of appropriate size are found, the threshold  $T_0$  is increased by a margin to  $T_1$ , and the search process above is repeated.

The number of blobs thus obtained are first merged when appropriate, based on adjacency requirements. Heuristics are then used to filter the remaining blobs and identify the one most likely to be part of the pupil of the eye. The heuristics which have been found useful include:

- 1) the number of detected pixels in each blob, roughly in the range (15, 100)
- 2) the position and value of the single darkest pixel in a blob

- 3) the ratio of the blob's height to its width, approximately in the range (0.33, 1.05)
- 4) the knowledge of the relative eye position in the face, and
- 5) the motion constraint that the eye movement is smooth and relatively small within two adjacent sampling frames.

The found pupil is then expanded proportionally, based on local information, to the size of 40 by 15 pixels to contain the cornea and the whole eye socket. Figure 3 shows an example of the segmented right eye image 300. (The right eye only is used in the embodiment of the present invention described herein but either eye could of course be used.)

The eye image segmentation approach described above is not very sensitive to changes in lighting conditions as long as the face is well lit (sometimes assisted by an ordinary desk lamp). It is not generally affected by the glasses the user is wearing either although, occasionally, strong reflections off the glasses and the appearance of the frame of the glasses in the segmented eye images due to the head moving away from the camera are problematic. They contribute a burst of noise which disrupts features in activation patterns (discussed below) to be sent to the purpose built neural network modelling system 125.

### *Histogram normalisation*

The segmented grayscale eye image, having a value between 0 and 255 for each pixel, is preprocessed by algorithms. For a real-time running system, the preprocessing algorithms should be simple, reliable and computationally not intensive. For instance, the algorithms might assume a value between -1.0 and 1.0 for each pixel. A neural network can then effectively discover the features inherent in the data and learn to associate these features and their distributions with the correct gaze points on the screen. Through adequate training, the network can then be endowed with the power to generalise to data that was not previously present. That is, it can use data learned in respect of similar scenarios and generate its own gaze point data from input data not previously encountered.

The histogram normalisation block 115 takes as input the individual 40 times 15 8-bit grayscale image and computes its histogram which is normally a unimodal

shape dominated by a main peak. An eye image whose histogram does not satisfy some desired requirements is rejected as a false segmentation. An example histogram is shown in Figure 4 where the lower and upper bounds,  $t_l = 36$  and  $t_u = 144$ , are found respectively.

5 In Figure 4, the vertical axis gives the number of pixels and the horizontal axis shows the grey levels over the range between 0 and 255, partitioned into 64 bins. The lower and upper bounds,  $t_l$  and  $t_u$  respectively, have grey scale values at 36 and 144. The region between the bounds is linearised (see below).

All the pixels within the range of 5% of the upper bound of the histogram are  
10 allocated a value 1.0 and all the pixels within the range of 5% of the lower bound of the histogram are allocated a value of -1.0. An arbitrary pixel ( $p$ ) falling within the remaining 90% of the histogram 400, between the bounds, assumes a linearised value  $t_p$  between -1.0 and 1.0. That is:

$$\Delta t = 0.05(t_u - t_l) \text{ and}$$

15 
$$t_p = -1 + 2(p - t_l - \Delta t) / (t_u - t_l - 2\Delta t)$$

Figure 5 shows the transfer function used for the normalisation procedure. In Figure 5,  $t_l$  and  $t_u$  are the lower and upper bounds of Figure 4.

The activation patterns thus generated, with associated properly coded output gaze points (discussed below), are ready for use in training a neural  
20 network. In real time operation mode, these patterns are inputs to the system for gaze prediction.

Figure 6 shows the same eye image as in Figure 3 after histogram normalisation. It illustrates that the contrast between important features (the eye socket, pupil, the reflection spot) has been significantly enhanced.

## 25 *Neural network*

Referring to Figure 1, the central part of the gaze tracking system is the neural network based modeller/tracker 125. There could be different strategies for choosing network topologies, training paradigms etc., subject to data format required, model complexities and real-time running constraints. In this example,  
30 the neural network is implemented in software and runs on workstation W although hardware net implementations can be used e.g. optical neural nets, as known in the art.

A suitable neural network is shown in Figure 7. This is a three-layer feedforward neural network with 600 input retina units 700, each receiving a normalised activation value from the segmented 40 x 15 eye image.

There are 16 hidden units 705, divided into two groups of 8 units each, and a split output layer 710 is introduced, deploying 50 and 40 units for describing respectively the horizontal and vertical positions of a screen gaze point. (The links as shown are fully connected.)

Figure 8 shows a matrix of 50 x 40 grids laid over a display screen 800 displayed on the display 101 of Figure 1A, for guiding the movements of a moving cursor and indicating the gaze position in order to collect the training data of eye image/gaze co-ordinate pairs for the neural network described.

As shown in Figure 8, this can correspond to dividing the display screen 800 uniformly into a rectilinear matrix of 50 by 40 grids, each sized about 23 by 22 pixels on the display. Depending on applications, the resolution of the grid matrix (50 times 40) can be increased or decreased. Also, if the viewing objects in an application are to appear in only part of the display screen 800, it suffices to collect the data (discussed below in the "Training Data Collection" Section) from this part of the screen and use them for training the model.

Referring again to Figure 7, the input units 700 are fully connected to the hidden layer units 705 which function as various feature detectors (further discussed below), but the connections between the hidden layer 705 and output layer units 710 follow the two separate groupings as indicated. Assuming the entire grid matrix defined is valid, the maximum number of connection weights (including biases) to be adapted amounts to:

$$t_{nc} = 601 \times (8 + 8) + 9 \times 50 + 9 \times 40 = 9616 + 810 = 10,426$$

All the hidden and output units 705, 710 assume a hyperbolic tangent transfer function of the form:

$$f(x) = (1 - e^{-x}) / (1 + e^{-x})$$

with  $f(x)$  having output values between -1.0 and 1.0.

It is interesting to note that using the whole eye image directly as the input to the neural network modeller actually provides a global holistic approach in contrast with the traditional explicitly feature-based approach.

### *Gaussian coding of output activations*

Given a grid matrix, 50 by 40 say, covering the whole display screen 101, as shown in Figure 8, the co-ordinates of an arbitrary gaze point 805 in this grid matrix can be a value between 0 and 49 along the "x" direction and between 0 and 39 along the "y" direction, with the origin being in the top left corner (0,0) of the screen. Instead of using the commonly seen "1 out of N" coding method for representing the desired activation pattern of a gaze point across the two groups of output units, respectively, a Gaussian shaped coding method has been adopted, based on earlier work published by Pomerleau (1993) in "Neural Network Perception for Mobile Robot Guidance": Kluwer Academic Publishing on autonomous vehicle guidance and by Baluja and Pomerleau (1994), published in "Non-intrusive gaze tracking using artificial neural networks": Technical Report CMU-CS-94-102, School of Computer Science, Carnegie Mellon University, on a similar gaze-tracking system.

It is generally agreed that the "1 out of N" coding method is more suitable for pattern classification tasks which require sharp definitive decision boundaries between different classes while the mapping function simulation task of embodiments of the present invention demand a gradual change in output representations when the data examples (eye appearance) in input data space exhibits slight difference. This preservation of topological relationships after data transformation (mapping) is the main concern in selecting an output coding mechanism.

Figure 9 shows a desired Gaussian shaped output activation pattern corresponding to the vertical position  $y = 15$  of a gaze point across the 40 output units 710. In the experiments discussed below, the Gaussian function used is of

the form  $G(n-n_0) = -1 + 2 \exp\left(\frac{-(n-n_0)^2}{10}\right)$  with the standard deviation  $\sigma = \sqrt{5}$ ,

which is depicted in Figure 9 at integer sampling positions. Paired (x, y) grid co-ordinates of a gaze point therefore give rise to two Gaussian shaped output activation patterns, taking values in the range between -1.0 and 1.0, one centred around the xth unit across the 50 output units for the horizontal axis and the other centred around the yth unit across the 40 output units for the vertical axis.

These two patterns concatenated together act as a desired output of the neural network system.

In decoding the outputs while testing the gaze tracking system, the Gaussian shaped activation pattern  $G(n-n_0)$  is moved across the output units for the x-coordinate by changing  $n_0$  from 0 to 49. A least-square fitting procedure is performed at each unit position to try to match the actual output activation pattern. The peak of the Gaussian shaped pattern that achieves the smallest error determines the horizontal position of the gaze point. Similarly, the vertical position of the gaze point across the 40 output units for the y-co-ordinates can be found.

10

### **Training and operation**

This section describes a means of collecting correct training data, the process of training a large neural network, analysing the significance of the learned connection weights and briefing the features regarding the real-time gaze tracking system.

15

### ***Training data collection***

For the gaze tracking system described above, one issue remaining is how to properly and automatically collect the training examples, or paired eye image/gaze co-ordinates, such that the neural network in question will be modelling the correct instead of a false mapping function. This allows the gaze tracking system to function properly and generalise to the real-time running situation. The following procedures are adopted:

20

- 1) The user is asked to visually track a blob cursor which travels along the grid matrix on the computer screen in one of the two predefined paths, obtaining horizontal/vertical zig-zag movements. At the outset, the travelling speed of the cursor can be adjusted to accommodate the acuity of the user's eye reaction time so that s(he) can faithfully and comfortably follow the moving cursor. The size of the blob or the resolution of the grid matrix (for indicating the position of the cursor) on the screen depends on the requirements of an envisaged application and the trade-off between running speed, system complexity and prediction accuracy. In the training phase, the smaller the blob is, the more images need to be collected in one session for the cursor to sweep through the entire screen grid matrix.

30



Accordingly, the neural network (described above) would have to make provisions for more output units to encode all the possible cursor positions.

2) At time  $t$ , when, the user visually tracks the moving cursor to a grid position  $(x, y)$ , the video camera, which keeps monitoring the user's head within its field of view, grabs a head image. From this image, a small patch of 40 by 15 pixels size containing appropriately the eye socket appearance is segmented. This eye image paired with the  $(x, y)$  co-ordinates of the travelling cursor forms one of the data examples for the neural network based gaze tracking system. The system has been designed principally to learn the complex mapping function from the appearance of the eye image to gaze position to the full extent of the computer screen.

3) One session of training images collection takes between 2 and 3 minutes. The cursor movement can be paused and resumed at a click of a mouse button. During the course of images collection, the user needs to satisfy some constraints for the current system to function properly. At the end of the recording session, the user can selectively download certain parts, or all of, the valid paired eye images/co-ordinates. The algorithm can detect automatically those unwanted images when eye blinks have occurred, and report a failure in capturing the eye image at that particular time and its associated gaze point. That is, the algorithm comprises a set of heuristics such that it can for instance learn a normal range of values and report failure when values fall outside the range.

4) To further remove some falsely segmented and recorded training examples, such as those corresponding to eye brows, nostrils or left eyes, the user can playback the downloaded image sequence at a selected speed and, if desired, visually examine and identify those noisy examples.

### ***Training of the neural network***

For the data collected in the manner above, and preprocessed and coded according to the discussion above, a backpropagation algorithm, see for instance Bishop, C. (1995): "Neural Network for Pattern Recognition" published by Oxford University Press, is used in order to train the neural network system. The cost function to be minimised is the usual summed squared error (SSE) which is subject to an evaluation criterion (for stopping purposes) called the average grid deviation

(AGD). The AGD measures the average difference between the current gaze predictions and the desired gaze positions for the training set, excluding a few wild cards due to the user's unexpected eye movements.

In the following, two strategies are discussed which can be used to train the large neural network. Starting with small random weights each having a value between -0.1 and 0.1, this first strategy consists of a fast search phase followed by a fine tuning phase.

In the first phase, the network is updated in terms of its weighting functions once for every few tens of training examples (typically between 10 and 30) which are drawn at random from the entire training data set. (It was found repeatedly that a training process taking examples in their original collection order would always fail to reach a satisfactory convergence state of the neural network, due to perhaps the network's catastrophic 'forgetting' factor.) A nominal learning rate  $r=0.4$  and a momentum factor  $m=0.5$  are adopted in training, which means that, for each connection weight  $w_i$ , the actual learning rate used for updating its value, varies, and is much smaller, equal to the nominal learning rate divided by the fan-in of the unit to which  $w_i$  is connected. A small offset  $\sigma = 0.05$  is added to the derivative of each unit's transfer function to speed up the learning process. This is especially useful when a unit's output approaches the saturation limits, either -1 or 1, of the hyperbolic tangent function. Besides, for each input training pattern random Gaussian noise is added, corresponding to 5% of the size of each retina input. This is particularly effective for overcoming the over-fitting problem in training a neural network and achieving better generalisation performance. In so doing, the neural network, albeit over ten thousand weights, would always approach a quite satisfactory solution after between 50 and 80 training epochs.

("Overfitting" in a neural network occurs when data which is simply noise is detected and learned as useful data by the system. This tends to occur when a network has too many nodes.)

In the second fine tuning phase, the network weights are updated once after presenting the whole training set. The nominal learning rate to use is proportionally much smaller than in the first phase, and a slightly smaller magnitude of Gaussian noise, around 3% of each retina input, is used. After about 30 epochs, the system can settle down to a very robust solution.

Figure 10 shows a trial learning result for user BA. The original data were collected in two horizontal and two vertical cursor running sessions, respectively. The cursor is confined to only travel within the top-right 40X30 area -- the application interested part of screen -- of the entire 50X40 sized grid matrix. So, each running session can provide at its maximum 1200 data examples. The total number of examples successfully collected for the four sessions is 3906, and the number of training examples used in obtaining the learning result of Figure 10 is 3000. The remaining 906 examples were used to examine the learning performance and to find the most appropriate stopping point. In this trial, the weights saved at the 60th epochs of training phase 1 are loaded for further refinement in phase 2. It can be seen that this overall strategy leads to rapid reduction in training error which then settles down to a stable status allowing for no further overfitting of the neural network.

#### 15 *Cross-validation*

Another strategy used is the cross validation technique. An independent validation set is set apart by randomly choosing original examples collected. During the course of training, the validation data set is involved to monitor the progress of the learning process in order to prevent the network from overfitting the training data. The learning process stops when the validation error starts to pick up or saturate after a previous general downwards trend. The weight set obtained at this point will be used to drive the real-time gaze tracking system. In practice, however, several trials with different initial weights are needed to find out the weight set with the smallest validation error, which is expected to provide better generalisation performance.

Figure 11 plots the curves for learning and its corresponding validation errors versus the number of training epochs in one trial of simulations of the neural network. (The ripples on the curves are due to the training scheme of using randomly chosen examples within each epoch and the way of updating the weights once for every ten examples instead of the whole training set of 2,800 examples.) Following the Section on data collection above, 1606 and 1635 valid paired examples have been collected, respectively, from observing the horizontal and vertical zig-zag cursor movement path, from which 206 and 235 randomly

chosen examples are correspondingly set apart to form a validation set of 441 examples. The remaining 2,800 examples constitute the final training data set.

It is clear from Figure 11 that as the training process is prolonged, the error of the training set continues decreasing while that of the validation set tends to reach a limit, on average, from around 100 training epochs onwards. The histogram of the weight set obtained at this point is given in Figure 12. Figure 12 shows the histogram of the neural network's connection weights (10,426 in total) after 100 training epochs. It follows a Gaussian distribution with an average value  $\mu = -0.0476$  and standard deviation  $\sigma = 0.8916$ . The width of the partition bin is 0.0412.

Figure 12 shows that the absolute majority of the connection weights lie in the range between -1 and 1. This is especially true for the connections between the input and hidden units, demonstrating the distributed nature (no dominant local impacts) of the neural network's collective responsibility. The weights with values beyond the range of (-1, 1) usually exist in the connections between the hidden and output unit, which combine, in a transformed and weighted way, the important features to determine the most appropriate response for the current input image.

#### *A real-time gaze tracking system*

The functions and procedures described in previous sections eventually lead to a trained gaze tracking system capable of operating in real-time. Referring to Figure 1, the switch S 120 is then connected to the real time running node R 135. Once the training (calibration) process is finished and an optimised weight set is loaded into the system, the gaze tracker will be ready to run. It constantly outputs the (x, y) gaze co-ordinates whenever it captures an eye image, or reports a failure when no eye is detected. As an indicator, a cursor is displayed on a reduced 50 x 40 grid window displayed on screen 101, to show where the user is looking on the entire computer screen.

The system works at about 20 Hz in its stand-alone mode on a SunUltra-1 Workstation. To gauge its performance, the average prediction accuracy on a separately collected test data set is about 1.5 degree, or around 0.5 inch apart on the computer screen. One can also test the system interactively by clicking the

mouse button on a grid point to highlight its position, then looking at it while clicking the mouse button again to show the system predicted gaze point.

If the calibration data (training examples) are collected uniformly over the entire screen grid, the system should perform equally well in all the areas. In practice, however, it has been observed that the gaze prediction error is actually distributed in a non-uniform fashion. In some parts of the screen it performs well as expected, but in other parts its prediction is relatively poor with occasionally some unexpected wild jumps. This problem may be relieved by introducing an offset table 103, after the neural network is fully trained, to adjust predictions in those badly performed areas in a real time running situation. The process of acquiring this offset table can be achieved based on the aforementioned interactive testing process, for example.

Another possible source of bad performance is the user's head orientation. It is possible to design the neural network to treat this as noise and therefore to discount or coalesce the data. Alternatively, it would be possible to improve the system's robustness by detecting and modelling also the user's head orientation, in addition to modelling the appearance of the eye. For this purpose, either the approach used by Stiefelhagen et al. (1996) and published in "Gaze Tracking for Multimodal human - computer interaction" (referenced above) can be applied with some modifications, or a second much smaller neural network with less than 10 inputs can be employed to learn the anthropomorphic features distribution against a gaze point.

The output of this head direction modeller can then be combined with the output of the current gaze tracker in a way to deliver a reliable and unambiguous indication of the gaze point, despite the head movements.

Training of the neural network, in view of the potential bottleneck it represents, will usually be done offline although the invention includes both online and offline training.

Thus, as previously mentioned, the neural network 125 shown in Figure 1B provides a stream of x, y, co-ordinate values when in the real time running mode indicative of the gaze direction of the user. This can be used to trigger operation of different functions on the workstation W. For example, by gazing at a

particular part of the screen, which displays an operating button display, the user can operate the display by looking at the button, in order to achieve similar functionality to a conventional mouse cursor. In order to achieve such functionality, post-processing techniques are performed, by a post-processing unit 5 135 shown in Figure 1B. Typically, a sliding window (not shown) is set up, which can be of variable size, to define an area of interest so that when the user gazes at it, the users gaze can be detected and used to trigger operation of a program or other workstation function. The post-processing is configured also to determine a so-called fixation i.e. when the user looks at the sliding window for a 10 predetermined time, so that a stream of x, y co-ordinate values fall into the window to produce a blob indicating that the output co-ordinates from the neural net show a serious intent by the user to operate the button, rather than the occurrence of a spurious glance. It will be appreciated that the response time to determine the serious intent of the user to gaze and achieve a fixation is an 15 inverse function of the response time of the system. Thus, response time can be traded off against accuracy. Similarly the size of the sliding window will affect the response time and accuracy.

These issues will now be discussed in more detail.

## 20 Adaptive processing of noisy data from the gaze tracker

It is possible, using embodiments of the present invention, to provide a user friendly interface by means of which the user can optimise external parameters for themselves. This is clearly of importance in a complex system in which it is not possible to map the effect of all the internal parameters on the external 25 parameters.

For instance, the user may favour fast response times over accuracy. By introducing a further network, for instance a Bayesian network, it is possible to develop a set of probabilistic rules which approximates a mapping of the effect of all the internal parameters on the external parameters. This is done by using a 30 fixed set of data points from the gaze which is fed into the system several times with various settings of the internal parameters. For instance, response time and

accuracy are measured for each fixation (training case). The network's conditional probabilities are then adapted to the examples.

External parameters which are of interest include response time, ie how fast the system detects a fixation, and accuracy in terms of the width (in a probabilistic sense) of the cluster. Internal parameters which can be involved include for instance, the size of the sliding window (n blobs in a cluster), the threshold for the line-fitting algorithm and the threshold for the horizontal.

A Bayesian network represents causal dependencies between variables. In this case, the variation in the internal parameters (thresholds and window size) will affect the system's behaviour (external parameters : response time and accuracy). In the learning phase, the system learns the conditional probabilities of the external parameters given the values of the internal parameters:

$$P(\text{response\_time} \mid \text{horiz\_threshold}, \text{fit\_threshold}, \text{window\_size})$$

$$P(\text{accuracy} \mid \text{horiz\_threshold}, \text{fit\_threshold}, \text{window\_size})$$

15 The Bayesian network is referenced 140 in Figure 1B.

After training, the network is ready to be used. In normal mode, the user can change the desired behaviour by setting one of the external parameters and the network propagates the influence of the new value backwards to the internal parameters which are thus adjusted. These adjusted values thus are used in the post-processor 135 shown in Figure 1B in order to control operation of the system according to the preferences set by the user.

## CLAIMS

1. A user interface, for use in making inputs to a data or communications system responsive to the user's eye, comprising:
  - 5 i) a scanning device for capturing a quantised image of an eye;
  - ii) a pupil image detector to detect a representation of the pupil of the eye in the quantised image;
  - iii) a display for a plurality of visual targets;
  - vi) a first learning device to relate at least one variable characteristic of said
  - 10 image of the eye to a selected one of said visual targets; and
  - vii) a second learning device, for relating external parameters apparent to a user of the system to parameters internal to the system.
2. An interface according to claim 1 which further comprises data
- 15 preprocessing means which can enhance the output of the scanning device for the purpose of the first learning device.
3. An interface according to claim 2 wherein the quantised image of the eye comprises an array of pixels with associated contrast information, and the data
- 20 preprocessing means comprises means to normalise said array and to allocate to each individual pixel thereof a contrast value selected from a set of discrete contrast values.
4. An interface according to any of the preceding claims wherein the learning
- 25 device comprises a neural network.
5. A user interface according to claim 4 wherein the neural network comprises a multiple layer feed forward network.



6. An interface according to claim 5 wherein the network includes a layer of input nodes, a layer of output nodes and a hidden layer of nodes providing a network of variable weight paths between the input and output nodes.
- 5 7. An interface according to claim 4, 5 or 6 operable in a training mode to train the neural network to correlate the direction of gaze of the eye of the user to said visual targets, and operable in a run mode to provide an output as a function of the direction of gaze of the user.
- 10 8. An interface according to claim 7 wherein the internal parameters are parameters that are a function of fixation of the gaze of the user at a particular region, and the external parameters include time taken to determine that a fixation has occurred and positional accuracy thereof.
- 15 9. An interface according to any preceding claim wherein the second learning device comprises a Bayesian net.
10. An interface according to any preceding claim including an offset table to introduce predetermined offsets as a function of the gaze direction of the user.
- 20 11. An interface according to any preceding claim wherein the scanning device comprises a video camera operable to capture a quantised image of the eye.
- 25 12. An interface according to claim 11 wherein the display comprises a workstation display screen and the video camera is mounted on the workstation.
13. A method of training a user interface that comprises:
- i) a scanning device for capturing a quantised image of an eye;
- 30 ii) a pupil image detector to detect a representation of the pupil of the eye in the quantised image;
- iii) a display for a plurality of visual targets;
- iv) a first learning device to relate at least one variable characteristic of said image of the eye to one of said visual targets; and

v) a second learning device, for relating external parameters apparent to a user of the system to parameters internal to the system;

the method comprising:

displaying training data on the display and training the first learning device to  
5 relate the variable characteristic of the image of the eye to the training data when the user gazes at the displayed training data.

14. A method according to claim 13 including training the second learning device to relate the external parameters apparent to the user of the system to the  
10 internal parameters.

15. A user interface for a computer workstation usable for videoconferencing, the interface being configured for use in making inputs to a data or communications system in response to movements of the user's eye, comprising:  
15 i) a tv videoconferencing camera to be mounted on the workstation for capturing a quantised image of an eye;  
ii) a pupil image detector to detect a representation of the pupil of the eye in the quantised image;  
iii) a workstation display for a plurality of visual targets; and  
20 viii) a neural net to relate at least one variable characteristic of said image of the eye to a selected one of said visual targets.

16. A gaze tracker including means for determining when a user achieves a gaze fixation on a target, comprising learning means for learning a relationship  
25 between response time and accuracy for achieving a fixation, and means responsive to a user's preference concerning the relationship for controlling signification of the fixation.

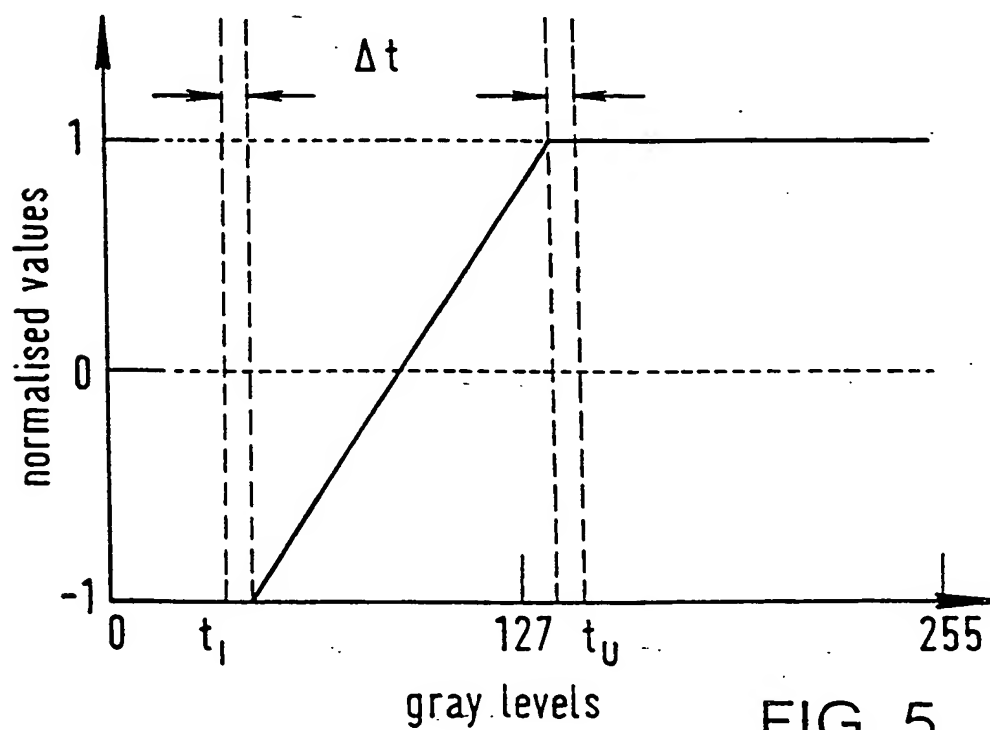
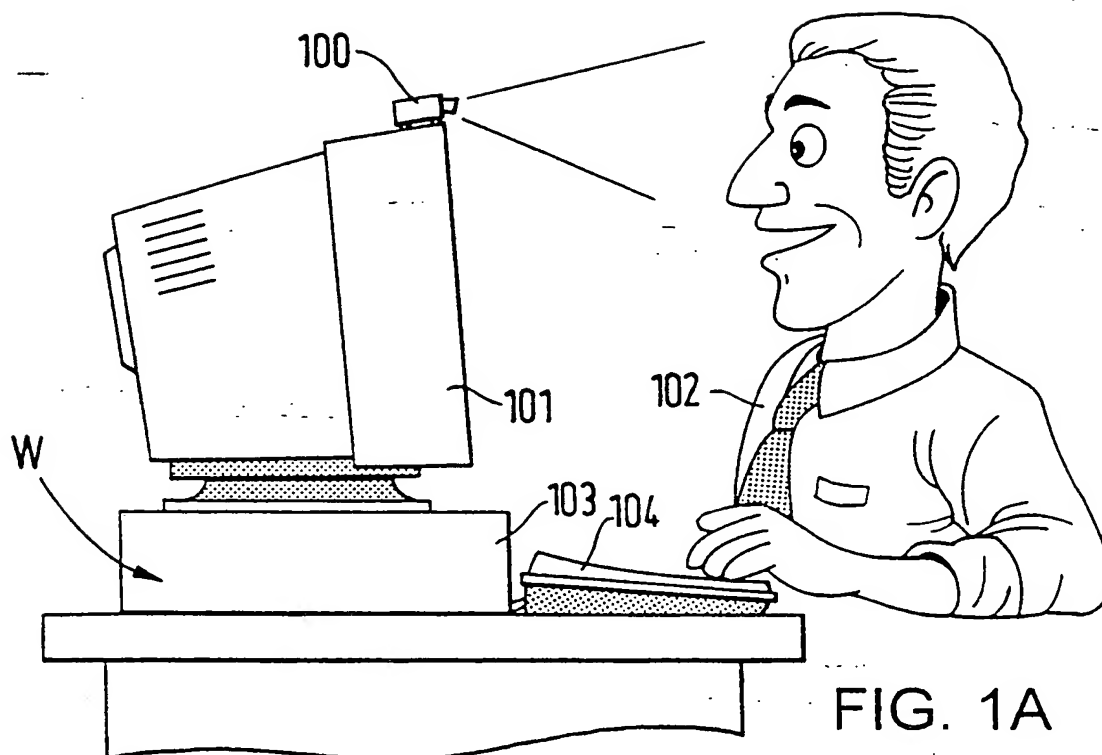
17. A gaze tracker according to claim 16 wherein the learning means comprises  
30 a Bayesian net.

18. A user interface substantially as hereinbefore described with reference to the accompanying drawings.

19. A method of training a user interface, substantially as hereinbefore described with reference to the accompanying drawings

5

1 / 7



2/7

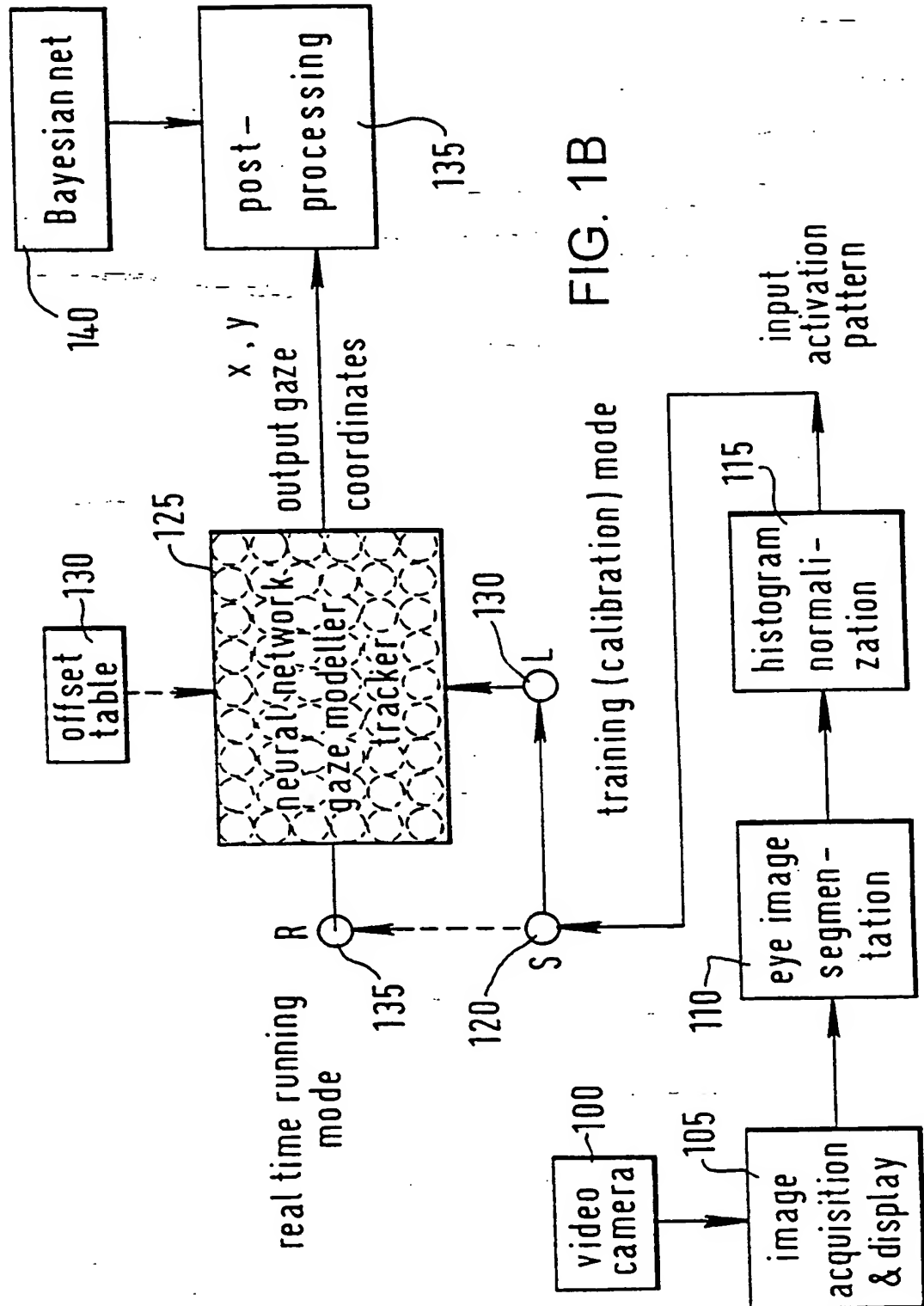


FIG. 1B

3 / 7

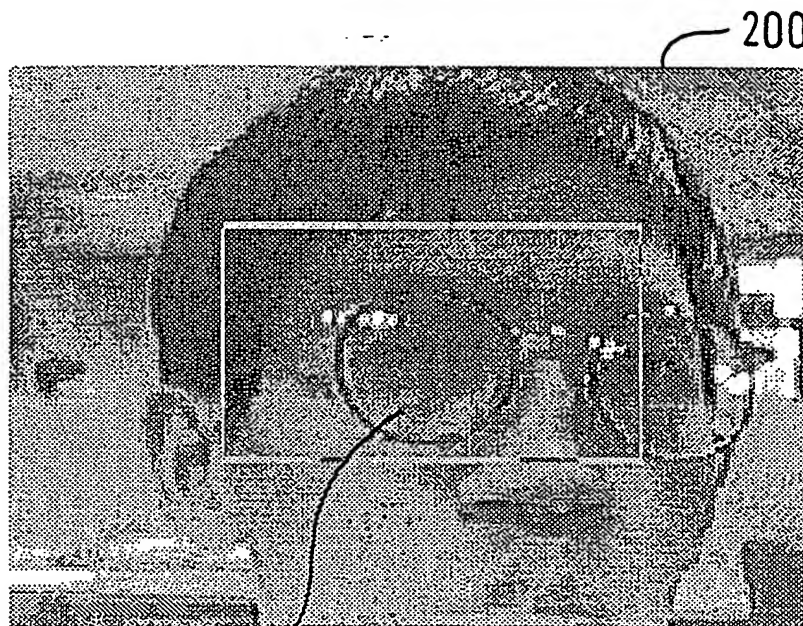


FIG. 2

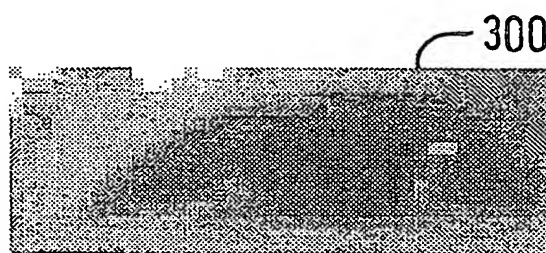
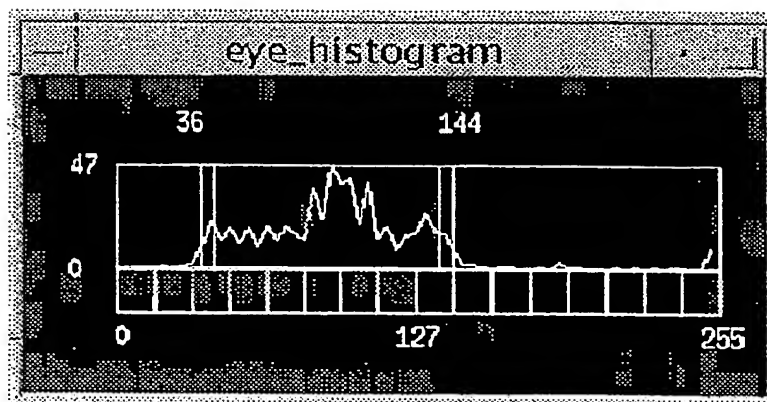


FIG. 3



$t_l$   $t_u$

FIG. 4

4 / 7



FIG. 6

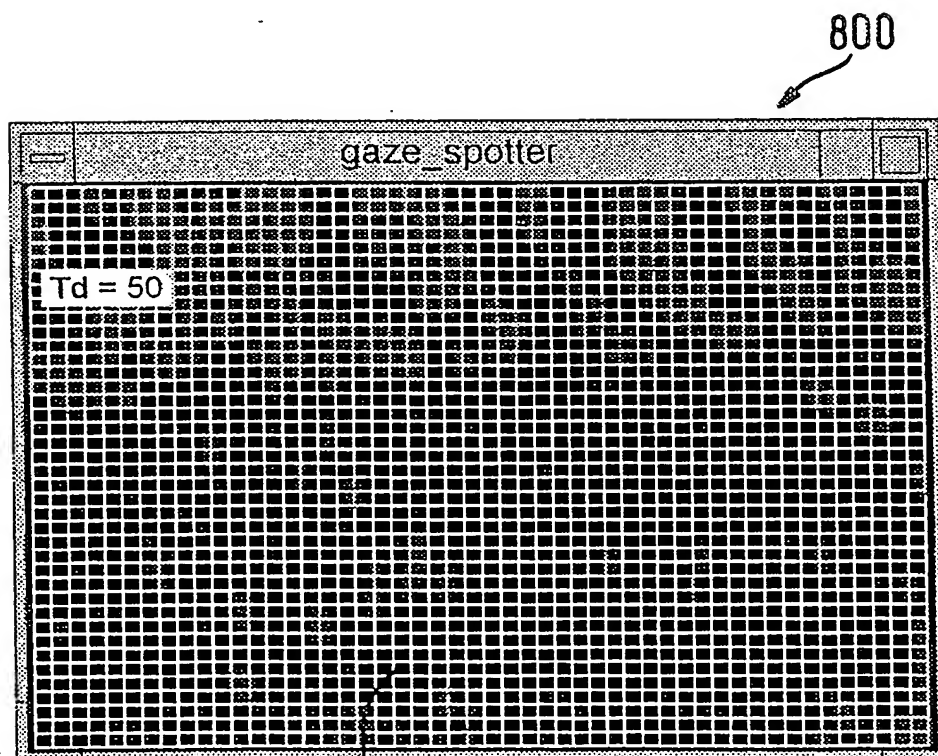
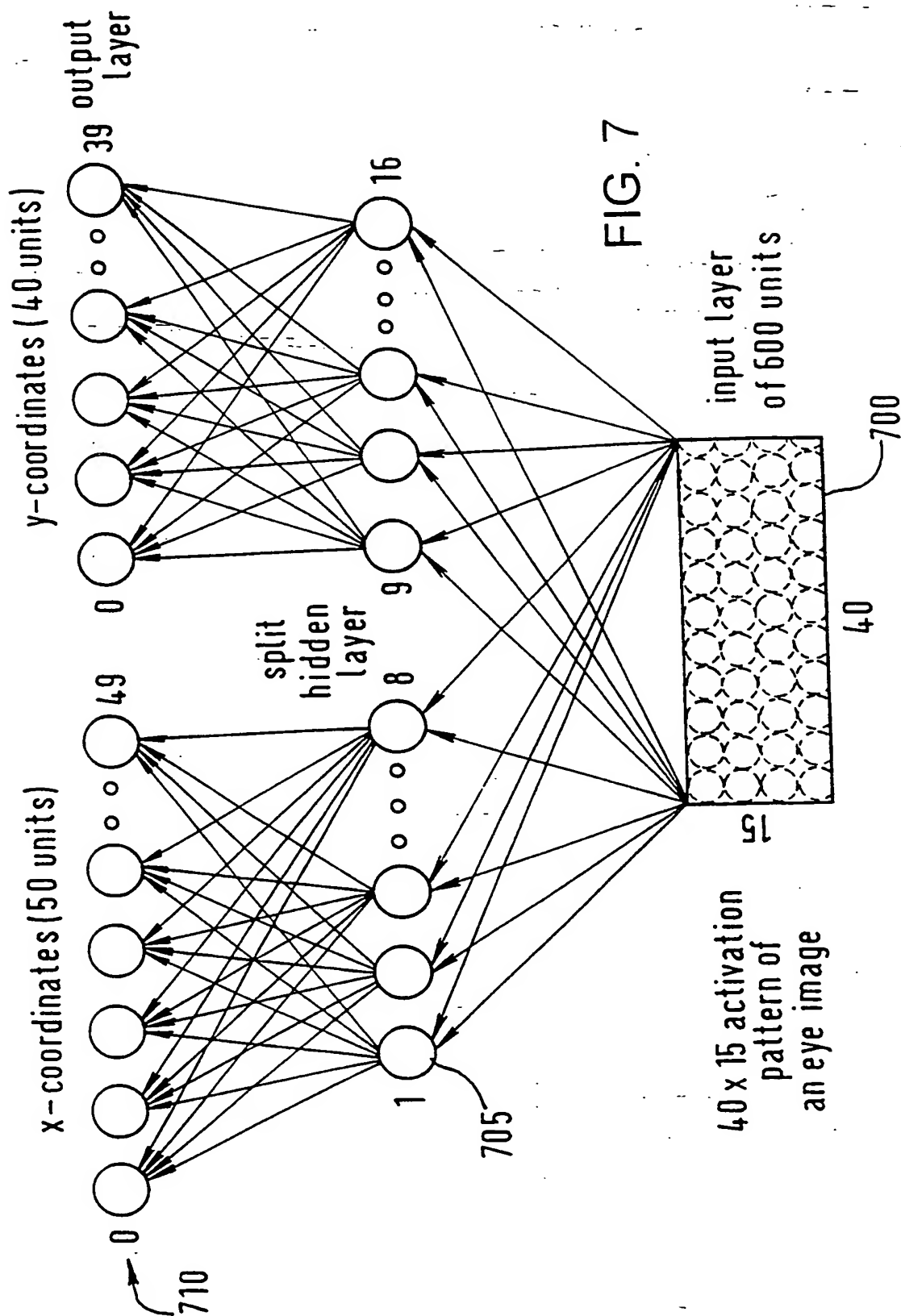


FIG. 8





6 / 7

FIG. 9

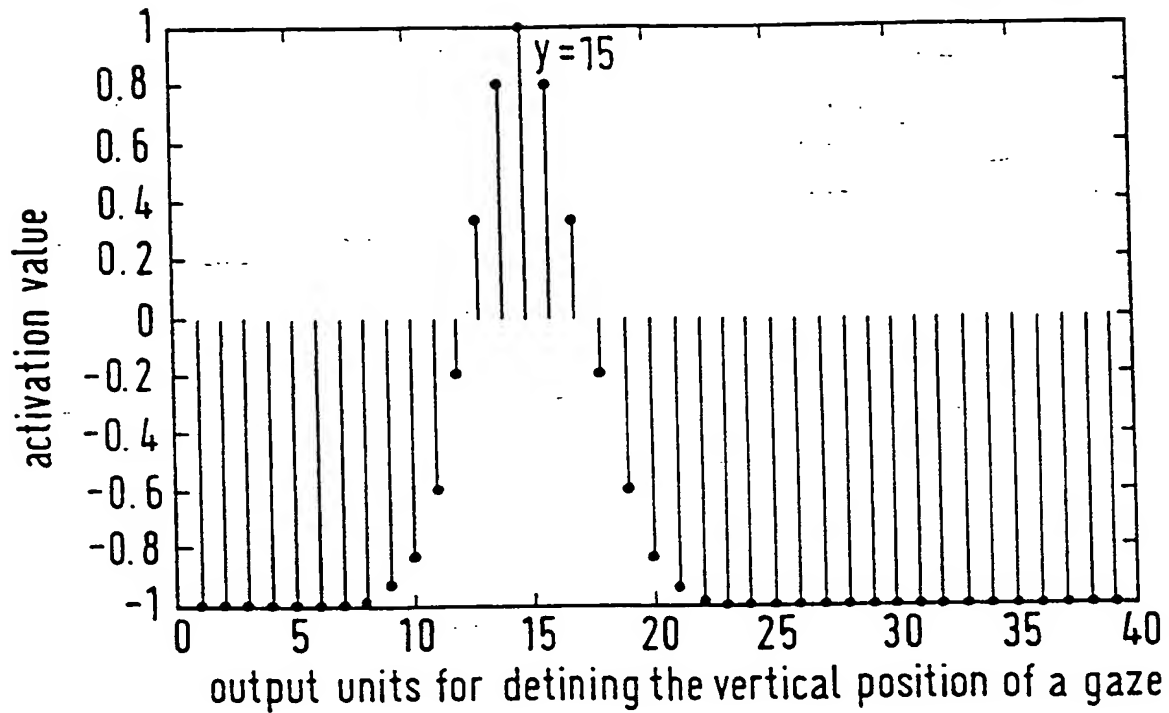
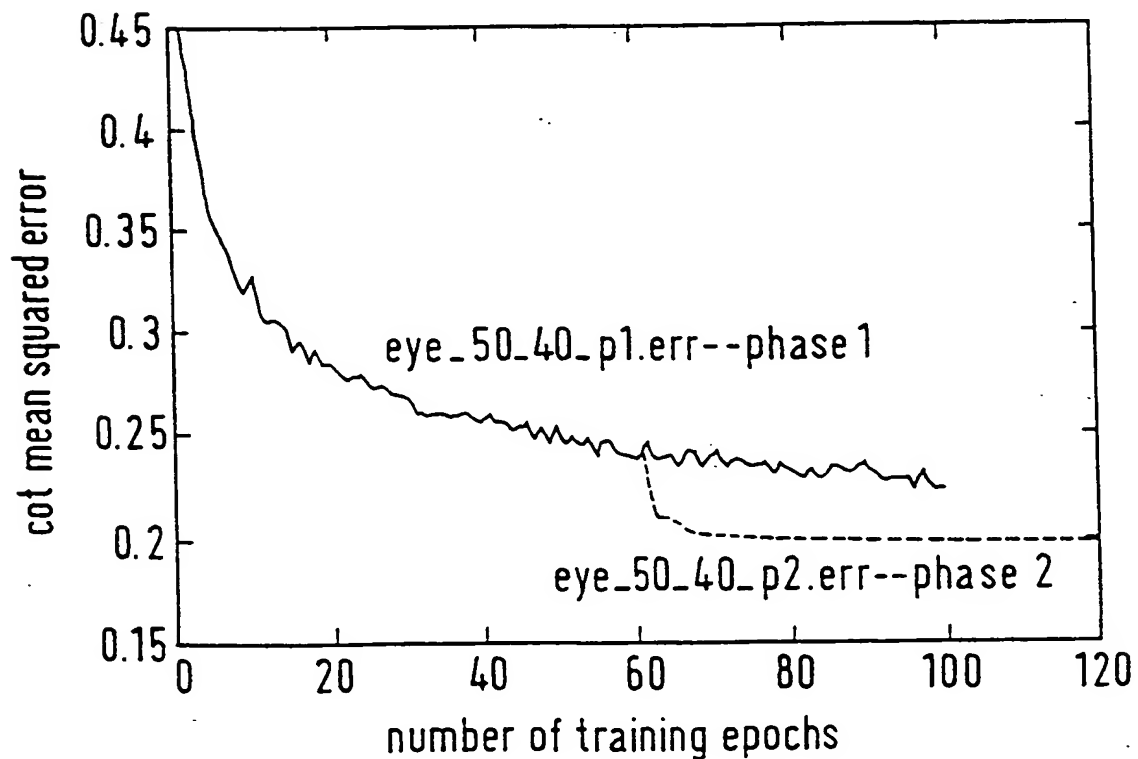


FIG. 10



7 / 7

FIG. 11

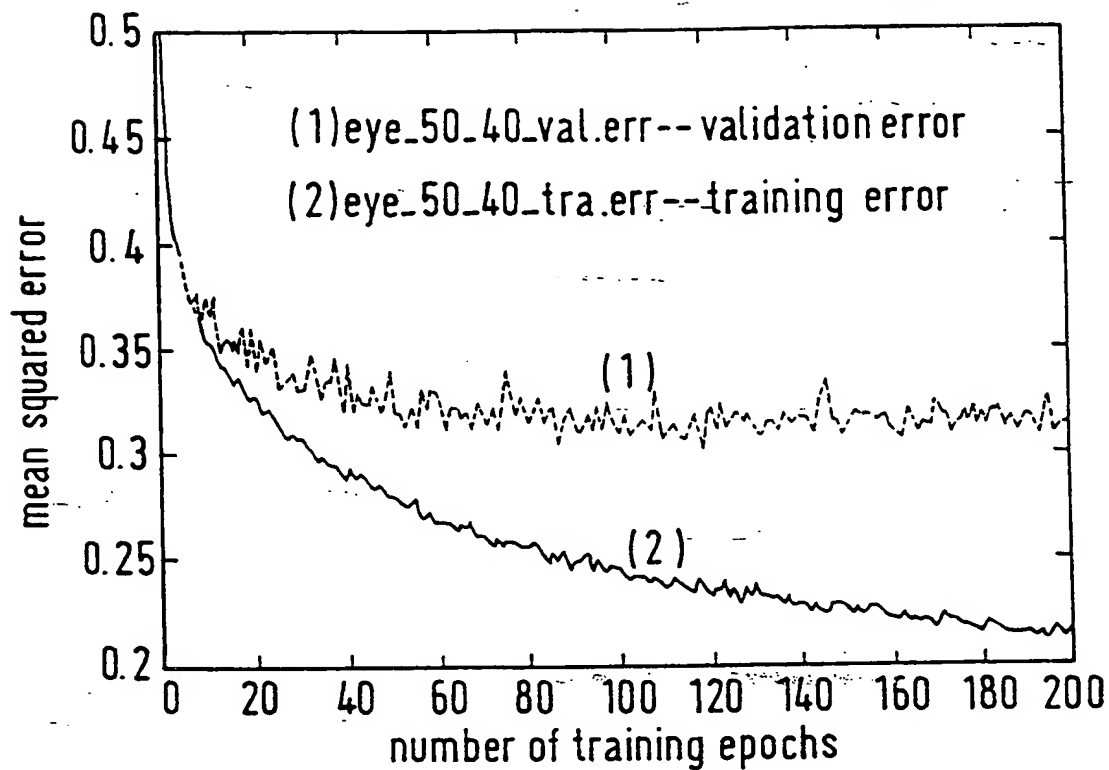
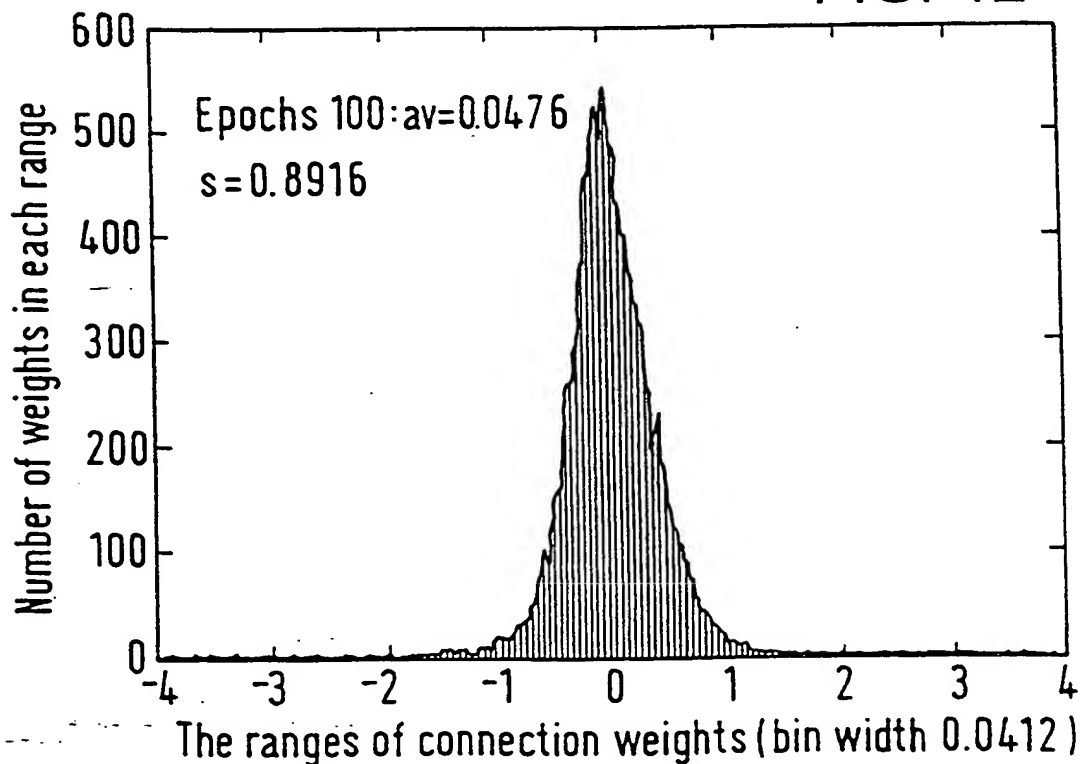


FIG. 12



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/03441

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 G06F3/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	TSUI K C ET AL: "Intelligent multi-modal systems" BT TECHNOLOGY JOURNAL, JULY 1998, BT LAB, UK, vol. 16, no. 3, pages 134-144, XP002083996 ISSN 1358-3948	15
A,P	see page 139, right-hand column, paragraph- 5 - page 141, right-hand column, paragraph 6; figures 8-11	1-14
Y	US 5 649 061 A (SMYTH CHRISTOPHER C) 15 July 1997	15
A	see column 3, line 39 - line 49 --- -/--	1,4-7,13



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

### \* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

8 January 1999

Date of mailing of the international search report

08. 02. 1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Durand, J

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/03441

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 471 542 A (RAGLAND RICHARD R) 28 November 1995	15
A	see column 3, line 33 - column 4, line 9 see column 11, line 15 - line 25; figures ---	1,13
A	US 5 481 622 A (GERHARDT LESTER A ET AL) 2 January 1996 see column 9, line 8 - line 30 see column 18, line 40 - line 57 see column 21, line 1 - line 26; claims 1-4 ---	1-3,11, 13,15,16
A	MITCHELL J L ET AL: "Human point-of-regard tracking using state space and modular neural network models" NEURAL NETWORKS FOR SIGNAL PROCESSING VI. PROCEEDINGS OF THE 1996 IEEE SIGNAL PROCESSING SOCIETY WORKSHOP (CAT. NO.96TH8205), NEURAL NETWORKS FOR SIGNAL PROCESSING VI. PROCEEDINGS OF THE 1996 IEEE SIGNAL PROCESSING SOCIETY WORKSHOP, KYOTO, JAPAN, 4 - 6 September 1996, pages 482-491, XP002089384 ISBN 0-7803-3550-3, 1996, New York, NY, USA, IEEE, USA see abstract see page 488 ---	1,13,15
A	WOLFE B ET AL: "A neural network approach to tracking eye position" INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION, 1997, ALEX PUBLISHING, USA, vol. 9, no. 1, 1997, pages 59-79, XP002089385 ISSN 1044-7318 see the whole document ---	1,13,15, 16
A	POMPLUN M ET AL: "An artificial neural network for high precision eye movement tracking" KI-94: ADVANCES IN ARTIFICIAL INTELLIGENCE. 18TH GERMAN ANNUAL CONFERENCE ON ARTIFICIAL INTELLIGENCE. PROCEEDINGS, KI-94: ADVANCES IN ARTIFICIAL INTELLIGENCE. 18TH GERMAN ANNUAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, SAARBRUCKEN, GERMANY, 18 - 23 September 1994, pages 63-69, XP002089386 ISBN 3-540-58467-6, 1994, Berlin, Germany, Springer-Verlag, Germany see the whole document -----	1,13,15, 16

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/GB 98/03441

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☒ Claims Nos.: 18,19  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
  
Claims referring only to the description and the drawings and not indicating clearly any special technical feature. (Rule 6.2(a) PCT)
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 98/03441

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5649061 A	15-07-1997	NONE	
US 5471542 A	28-11-1995	NONE	
US 5481622 A	02-01-1996	NONE	

Form PCT/ISA/210 (patent family annex) (July 1992)